

ORIGINAL RESEARCH

Signals of positive selection in mitochondrial protein-coding genes of woolly mammoth: Adaptation to extreme environments?

Jacob Njaramba Ngatia¹ | Tian Ming Lan^{2,3,4} | Thi Dao Dinh¹ | Le Zhang¹ |
Ahmed Khalid Ahmed¹  | Yan Chun Xu^{1,5,6} 

¹College of Wildlife Resources, Northeast Forestry University, Harbin, China

²BGI-Shenzhen, Shenzhen, China

³Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, Copenhagen, Denmark

⁴China National Genebank, BGI-Shenzhen, Shenzhen, China

⁵State Forestry and Grassland Administration Research Center of Engineering Technology for Wildlife Conservation and Utilization, Harbin, China

⁶State Forestry and Grassland Administration Detecting Centre of Wildlife, Harbin, China

Correspondence

Yan Chun Xu, College of Wildlife Resources, Northeast Forestry University, Harbin 150040, Heilongjiang, China.
Email: xu_daniel@163.com

Funding information

Fundamental Research Funds for the Central Universities of China, Grant/Award Number: 2572017PZ02; Wildlife Conservation and Management Project of State Forestry and Grassland Administration of China, Grant/Award Number: 201706; Guangdong Provincial Academician Workstation of BGI Synthetic Genomics, BGI-Shenzhen, Guangdong, China; The Guangdong Provincial Key Laboratory of Genome Read and Write, Grant/Award Number: 2017B030301011

Abstract

The mammoths originated in warm and equatorial Africa and later colonized cold and high-latitude environments. Studies on nuclear genes suggest that woolly mammoth had evolved genetic variations involved in processes relevant to cold tolerance, including lipid metabolism and thermogenesis, and adaptation to extremely varied light and darkness cycles. The mitochondria is a major regulator of cellular energy metabolism, thus the mitogenome of mammoths may also exhibit adaptive evolution. However, little is yet known in this regard. In this study, we analyzed mitochondrial protein-coding genes (MPCGs) sequences of 75 broadly distributed woolly mammoths (*Mammuthus primigenius*) to test for signatures of positive selection. Results showed that a total of eleven amino acid sites in six genes, namely ND1, ND4, ND5, ND6, CYTB, and ATP6, displayed strong evidence of positive selection. Two sites were located in close proximity to proton-translocation channels in mitochondrial complex I. Biochemical and homology protein structure modeling analyses demonstrated that five amino acid substitutions in ND1, ND5, and ND6 might have influenced the performance of protein–protein interaction among subunits of complex I, and three substitutions in CYTB and ATP6 might have influenced the performance of metabolic regulatory chain. These findings suggest metabolic adaptations in the mitogenome of woolly mammoths in relation to extreme environments and provide a basis for further tests on the significance of the variations on other systems.

KEYWORDS

Mammuthus primigenius, mitochondrial genome, mitolineages, positive selection, woolly mammoth

1 | INTRODUCTION

The genera *mammuthus* is believed to have originated in warm and equatorial Africa around 6.7–7.6 million years ago (Rohland et al., 2007) and later colonized cold and high-latitude environments in the arctic (Lister & Bahn, 2007). The woolly mammoth, *Mammuthus primigenius*, was the last living species of this genera, whose final population on Wrangel Island survived until about 4,000 years ago (Palkopoulou et al., 2015). Genomic studies have suggested three major clades in woolly mammoths (Chang et al., 2017; Palkopoulou et al., 2013). Clade 1 includes three mitochondrial haplogroups, namely C, D, and E, which occurred in North America and Eurasia. Clade 2/A occurred in Asia, but was geographically restricted to northern Siberia and western Beringia, where it coexisted in sympatry with Clade 1 (haplogroups D and E) until its extinction (Barnes et al., 2007; Chang et al., 2017; Debruyne et al., 2008). Clade 3 which includes haplogroups B1 and B2 were found in Alaska/Yukon of North America and Europe, respectively (Chang et al., 2017; Palkopoulou et al., 2013).

The woolly mammoths were suggested to have lived in high-latitude environments with extreme climatic (cold) and dry conditions (MacDonald et al., 2012). Studies have suggested that nuclear genes of woolly mammoth had evolved genetic variations involved in processes relevant to cold tolerance, including lipid metabolism and thermogenesis, and adaptation to extremely varied light and darkness cycles (Smith, Kawash, Karaikos, Biluck, & Grigoriev, 2017). However, the most essential role in generating energy in eukaryotic cells is played by the mitochondria where up to 95 percent of energy is generated through oxidative phosphorylation (OXPHOS) system (Levin, Blumberg, Barshad, & Mishmar, 2014). Different mitochondrial variants may have different effects on cellular and organismal metabolic performance (Ballard & Melvin, 2010). Thus, it could be expected that the mitochondrial genome of the woolly mammoth would have evolved in association with nuclear genome in adaptation to such extreme environments. Up to the present, little is known regarding adaptive evolution of their mitochondrial genes.

In vertebrates, 13 mitochondrial protein-coding genes (MPCGs) are involved in phosphorylation process, which includes seven subunits of NADH dehydrogenase (ND), one subunit of cytochrome *b* (cyt *b*), two subunits of ATP synthase (ATP), and three subunits of cytochrome *c* oxidase (COX) (Wallace, 2007). These subunits are fundamental in electron and proton transportation, and respiratory regulation of the mitochondria (Bai, Shakeley, & Attardi, 2000). In addition, the subunits have close association with those encoded by nuclear genome in the OXPHOS pathway (Rand, Haney, & Fry, 2004). Thus, OXPHOS functions can be influenced by mutations occurring in either genome, and mitogenome products are especially important to directly mediate proton pumping across the mitochondrial membrane (Gu et al., 2016). These findings suggest that co-evolution between the two genomes has been favored by natural selection to promote optimal maintenance of metabolic functions.

The impact of any amino acid substitution can differ significantly depending on its position in the protein structure. Therefore,

site-specific mutations that do not indicate changes in protein structure can influence function, for example changes in the binding sites. Amino acid substitutions in the mitochondrial subunits can influence the efficiency in the electron transport chain system, that can hinder or improve the OXPHOS process (Fonseca, Johnson, O'Brien, Ramos, & Antunes, 2008) and can be linked to environmental adaptations (Almeida, Maldonado, Vasconcelos, & Antunes, 2015). Adaptive changes observed in the MPCGs have been associated with environmental conditions, such as latitudinal clines (Stager, Cerasale, Dor, Winkler, & Cheviron, 2014), anoxic conditions (Tomasco & Lessa, 2011), and altitude (Yu, Wang, Ting, & Zhang, 2011).

Here, we compared comprehensive dataset of mitogenome sequences of woolly mammoth's to identify positively selected sites in the MPCGs, assessed the biochemical changes in amino acids and applied structure-based homology comparison models to predict the effect of these changes in the protein structure and functioning of the OXPHOS pathway.

2 | METHODS

2.1 | Sequence collection

Unpublished whole genome sequence data of six woolly mammoth (*Mammuthus primigenius*) samples were downloaded from NCBI GenBank (<https://www.ncbi.nlm.nih.gov>: Accession number ERP111819) or CNSA (<https://db.cngb.org/cnsa>: Accession number CNP0000209), and subsequently used to generate complete mitochondrial genomes. Raw reads were filtered using AdapterRemoval (Lindgreen, 2012) with the minimum base quality set to 15 and the minimum read lengths set to 30. The first 2 bases and the last 3 bases at the ends of each read were trimmed to prevent DNA damage-induced errors because deamination is more likely to occur at the ends of ancient DNA fragments. The C-to-T and G-to-A changes that occurred at the first 10 and last 10 bases of each read were also removed to further avoid potentially incorrect bases. Filtered reads were aligned to the reference genome of African elephant (loxAfr4, <ftp://ftp.broadinstitute.org/distribution/assemblies/mammals/elephant/loxAfr4/>) with the mitochondrial genome being replaced by woolly mammoth mitogenome (Acc No. DQ188829.2) using BWA (Li & Durbin, 2009) with the *aln* mode using parameters: -l 30 -k 2 -t 4. The multi-mapped reads were removed to obtain uniquely mapped reads. SAMtools (Li et al., 2009) was then used to remove the unmapped reads as well as reads with a mapping quality less than 1. We finally drew a random sequence for each site on the reference mitogenome to retrieve the mitochondrial genome for each of the six samples (Yang et al., 2017). The retrieved mitochondrial genome sequence of one of the six samples (N6) was found to contain numerous sites with gaps. Thus, it was not included in further analyses. Seventy additional mitogenome sequences (Chang et al., 2017; Enk et al., 2016; Fellows Yates et al., 2017; Gilbert et al., 2008, 2007; Krause et al., 2006; Rogaev et al., 2006) were downloaded from GenBank and included in the analysis (Table S1). These samples were chosen based on the previously

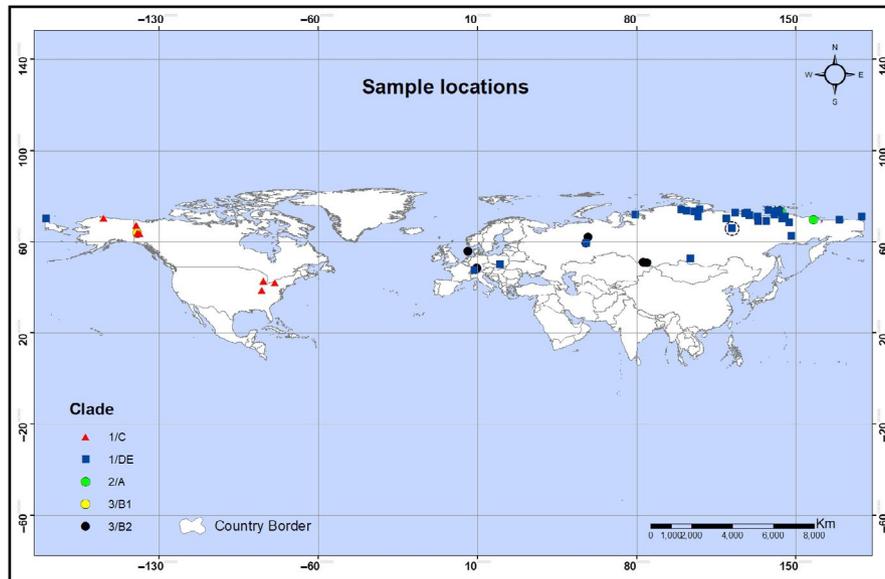


FIGURE 1 Sites locations of woolly mammoth mitochondrial genome samples included in this study. The samples are color-coded by clade. The location of the five samples we collected and used for this study is shown as a blue square enclosed with a dotted circle, of which the latter represents the approximate diameter for their geographic range

published mitolineage (clade) information, geographic origin, and the completeness of the genome sequences. The selected mitogenome sequences were complete or with relatively fewer gaps in the mitochondrial protein-coding genes region and were considered more suitable for subsequent analyses. The geographic origins of the 75 samples are shown in Figure 1. Previously published mitogenome sequences of *Mammuthus columbi* (GenBank accession No JF912199) and *Mammuthus jeffersonii* (GenBank accession No KX027559) were also included in this study.

2.2 | Alignment and phylogenetic analyses

Thirteen protein-coding gene (ND1- 6, ND4L, ATP6, ATP8, COX1, COXII, COXIII, CYTB) sequences were extracted from complete mitogenome and concatenated using Geneious Pro 7.0.6 (Kearse et al., 2012). Nucleotide sequences were aligned using ClustalW in MEGA 5 (Tamura et al., 2011), where the Asian elephant (GenBank accession No EF588275) was used as an outgroup. Alignments and subsequent analyses including positive selection were performed independently for each gene. But for phylogenetic analyses, concatenated sequences of the coding genes were considered. Here, Bayesian analysis (MB) was done using MrBayes (Ronquist & Huelsenbeck, 2003). The best-fit model of evolution was identified using jModelTest (Darriba, Taboada, Doallo, & Posada, 2012), through the Akaike information criterion (AIC) (Bozdogan, 1987). But because the selected model (TVM + G) is not implemented by MrBayes, it was substituted by GTR model (Lecocq et al., 2013). Bayesian MCMC (Markov chain Monte Carlo) was run for 2 million generations and sampling every 1,000 generations. The phylogenetic tree was summarized in MrBayes after removing the first 25% of the tree as burn-in. The consensus tree was then visualized using Fig Tree. With respect to selection analyses, nucleotide sequences

for each gene were separately aligned against the reference sequence Acc. No EF588275. Maximum likelihood (ML) trees included in TreeSAAP analysis were constructed using models of evolution identified as mentioned above (Table 1).

2.3 | Selection analyses

The widely used method for detecting selection in protein-coding genes is by estimating ω , the nonsynonymous/synonymous substitution rates ratio (d_N/d_S) (Yang, 1998). However, this approach for identifying genetic adaptation is mostly biased against even the proteins that are moderately conservative, since it does not accommodate the likelihood that adaptation may result from a few amino acid changes.

TABLE 1 The selected models of evolution for each protein-coding gene as inferred by jModelTest

Gene	Model
ND1	TPm3uf + I
ND2	TIM3 + I
ND3	HKY + G
ND4L	TPM2uf
ND4	HKY + I
ND5	HKY
ND6	HKY + I
CYTB	HKY + G
COX I	TIM2 + G
COX II	TrN
COX III	TPM1uf + I
ATP6	TPM3uf
ATP8	HKY

Therefore, significant changes in physicochemical properties among protein residues in a phylogenetic tree were detected using TreeSAAP, which compares observed distribution of physicochemical changes deduced from a phylogenetic tree with expected distributions based on an assumption that completely random amino acid substitutions are expected under selective neutrality conditions (Woolley, Johnson, Smith, Crandall, & McClellan, 2003). To reduce false positives in TreeSAAP analysis, we included only 20 amino acid properties showing >85% accuracy of identifying positive selection, set a sliding window equal to 15 codons, and only considered amino acid changes with radical effects on physicochemical properties (magnitude categories ≥ 6 and $p < 0.001$). Significant positive Z-scores imply that higher magnitude nonsynonymous mutations are more frequent than expected under neutrality, suggesting change due to positive selection (McClellan & Ellison, 2010; Woolley et al., 2003). Significant nonsynonymous

mutations found to occur at the clade or haplogroup level were considered as candidates of positive selection in this study.

2.4 | Protein modeling and structural analyses

In order to test the effect of the biochemical changes of the selected sites on the protein structure of *M. primigenius*, the protein sequences of each subunit of complex I were aligned with sequences of homologous subunits in GenBank (<https://www.ncbi.nlm.nih.gov/>) protein database. Protein homology models of individual subunits were then predicted using the automated mode program (<http://swissmodel.expasy.org/workspace/index>) of the SWISS-MODEL server using the best hit templates. The best hit homology templates for each subunit were as follows: ND1 (PDB: 5lnk.1.H) and ND5 (PDB: 5lnk.1.M), both from *Ovis aries*: ND2 (PDB: 5gup.34.A),

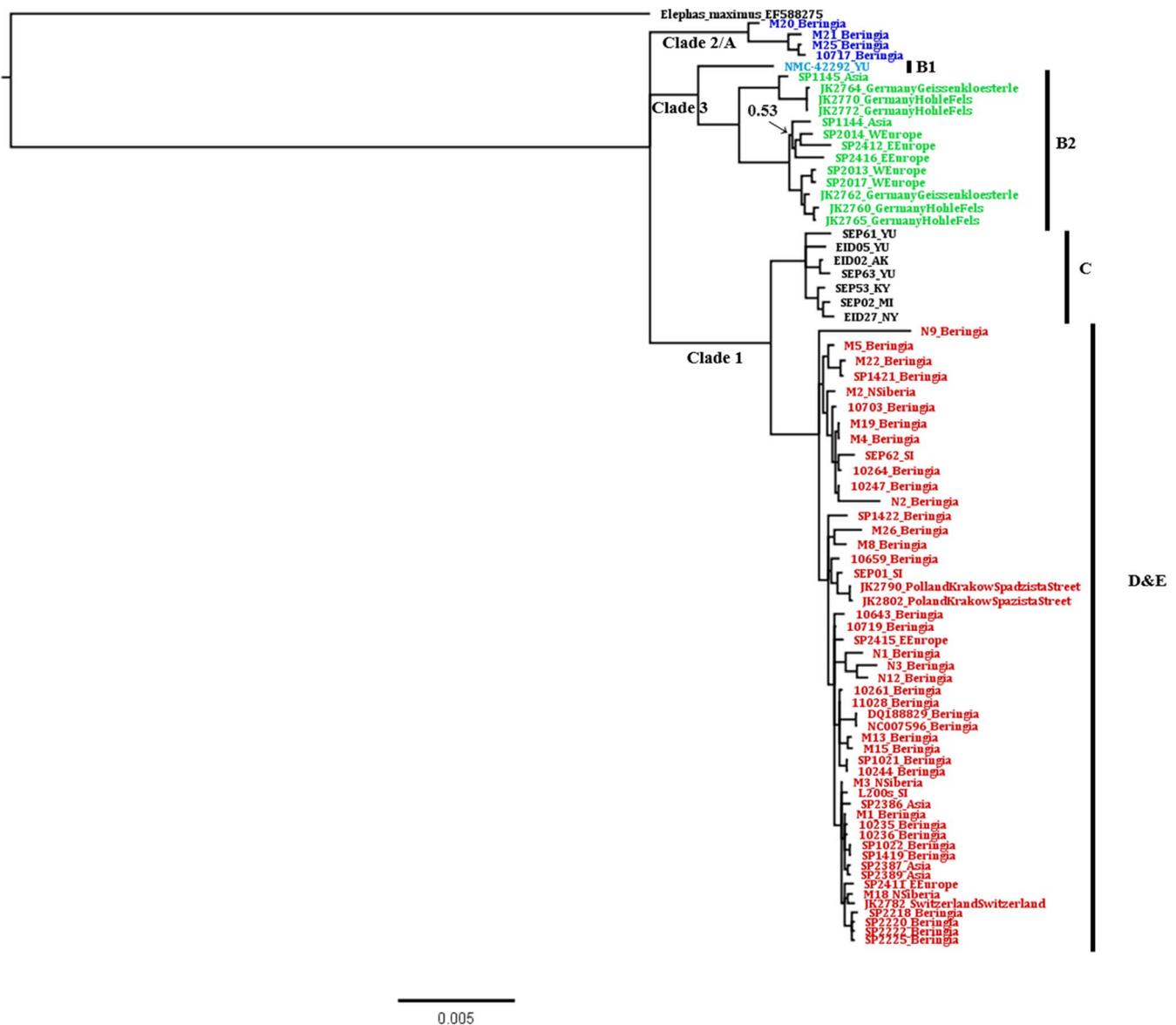


FIGURE 2 Rooted phylogeny of *Mammuthus primigenius* based on mtDNA protein-coding sequences estimated by Bayesian inference. All the nodes are well supported (PP > 0.9) unless indicated otherwise; mtDNA clades and the respective haplogroups labeled on the tree branches and on the right side of the tree respectively; the taxa names and codes are the individual sequence ID codes

TABLE 2 Nonsynonymous substitutions influencing changes in biochemical properties as inferred by TreeSAAP

Gene (codon)	Codon changes	Amino acid replacement	Physicochemical properties						
			N_s	pK'	Br	Ra	Hp	Ht	
ND1(6)	ATC-GTC	Ile-Val		↑8					
ND1(73)	ATT-GTT	Ile-Val		↑8					
ND2 (282)	ATG-ATA	Met- Ile		↓8					
ND4 (87)	GAA-GGA	Glu-Gly		↓8					
ND5 (16)	ATC-ACC	Ile-Thr	↓6	↑6	↓6	↓7	↓6	↓7	
ND5 (20)	ACC-ATC	Thr-Ile	↑6	↓6	↑6	↑7	↑6	↑7	
ND5 (51)	ATT-ACT	Ile-Thr	↓6	↑6	↓6	↓7	↓6	↓7	
ND5 (186)	ATG-ATA	Met-Ile		↓8					
ND5 (191)	ACC-ATC	Thr-Ile	↑6	↓6	↑6	↑7	↑6	↑7	
ND5 (512)	ACA-ATA	Thr-Ile	↑6	↑6	↑6	↑7	↑6	↑7	
ND6 (4)	ATT-GTT	Ile-Val		↑8					
CytB (189)	ATG-ATA	Met-Ile		↓8					
CytB (300)	GTC-ATC	Val-Ile		↓8					
CytB (371)	ATA-ATG	Ile-Met		↑8					
ATP6 (110)	GTC-ATC	Val-Ile		↓8					
COX1(190)	GTT-ATT	Val-Ile		↓8					

Note: The codon number corresponds to the coding alignment of each gene as applied in this study. The values in the physicochemical properties columns represent the category of radical change in each property, where (↑) shows an increase in magnitude, while (↓) shows a decrease. Amino acid change was significant at categories 6–8.

N_s = number of surrounding residues; pK' = equilibrium constant (ionization of COOH); Br = buriedness; Ra = solvent accessible reduction ratio; Hp = surrounding hydrophobicity; Ht = thermodynamic transfer hydrophobicity.

ND3 (PDB: 35gup.35.A), ND4L (PDB: 5gup.36.A), and ND4 (PDB: 5gup.42.A) that were all from *Sus scrofa*; and ND6 (PDB: 5lc5.1.J) from *Bos Taurus*. An overall structure of the *M. primigenius* complex I was then generated by structurally aligning individual subunits to corresponding complex I template from *Ovis aries* (PDB:5lnk) using Chimera software (Pettersen et al., 2004). The modeled complex I structure was refined by adjusting the side chains using FoldX software (Schymkowitz et al., 2005). After modeling complex I structure, solvent accessibility area (ASA) of each residue under positive selection was computed using GETAREA program (Fraczkiewicz & Braun, 1998) and determined whether they were located on protein binding sites using PRODIGY (Vangone & Bonvin, 2015; Xue, Rodrigues, Kastritis, Bonvin, & Vangone, 2016). Residue locations on the protein were established using TMHMH v2.0 (2001). Intraprotein cavities and pathways were examined with program CAVER (Petřek et al., 2006). Functionally important residues were determined using ConSurf (Glaser et al., 2003). Graphic modifications and presentations were finally depicted using PyMOL (DeLano, 2002).

3 | RESULTS

3.1 | Phylogenetic analyses

Based on the concatenated sequences of 13 protein-coding genes (11,412 bp), a phylogenetic tree constructed using MrBayes showed

distinct woolly mammoth's mitolineages and was congruent with those previously constructed using complete mitochondrial genomes (Chang et al., 2017; Gilbert et al., 2008). Most nodes were well supported (posterior probability (PP) > 0.9), with each clade showing PP of 1. All the sequences prepared in this study (N1, N2, N3, N9, and N12) were found to belong to Clade 1 haplogroups D and E (Figure 2).

3.2 | Evidence of positive selection in mitochondrial genes

TreeSAAP, which is highly sensitive to even the most conservative proteins, detected sixteen significant ($p < 0.001$) and radical (categories 6–8) physicochemical changes in amino acid properties in eight of the thirteen protein-coding genes (Table 2). The nonsynonymous mutations identified in codons ND1, ND2, ND4, ND6, Cyt *b*, ATP6, and COX1 showed significant changes in equilibrium constant (ionization of COOH) (pK'). In ND5, the mutation influenced radical changes in average number of surrounding residues (N_s), equilibrium constant (ionization of COOH) (pK'), buriedness (Br), solvent accessible reduction ratio (Ra), surrounding hydrophobicity (Hp), and thermodynamic transfer hydrophobicity (Ht) (Table 2). Changes in pK' imply changes in ionization properties of the amino acids while changes in N_s , Br, Ra, Hp, and Ht indicate changes in hydrophilicity or compactness in the surrounding region of the mutated residue (Cozzone, 2010).

TABLE 3 Calculations of protein structures prediction

Complex	Gene	Parent template	Species	% Similarity to Parent structure	Residues predicted	Reference
I	ND1	5lnk.1.H	<i>Ovis aries</i>	78.55	1-317	Fiedorczuk et al. (2016)
I	ND2	5gup.34.A	<i>Sus scrofa</i>	63.11	1-347	Wu, Gu, Guo, Huang, and Yang (2016)
I	ND3	5gup.35.A	<i>Sus scrofa</i>	78.95	1-114	Wu et al. (2016)
I	ND4L	5gup.36.A	<i>Sus scrofa</i>	73.47	1-97	Wu et al. (2016)
I	ND4	5gup.42.A	<i>Sus scrofa</i>	74.2	1-438	Wu et al. (2016)
I	ND5	5lnk.1.M	<i>Ovis aries</i>	69.97	1-597	Fiedorczuk et al. (2016)
I	ND6	5lc5.1.J	<i>Bos taurus</i>	57.89	1-171	Zhu, Vinothkumar, and Hirst (2016)

Note: The parent template and species columns show the best hit parent structures used in predicting the individual protein chains, and the corresponding species, respectively.

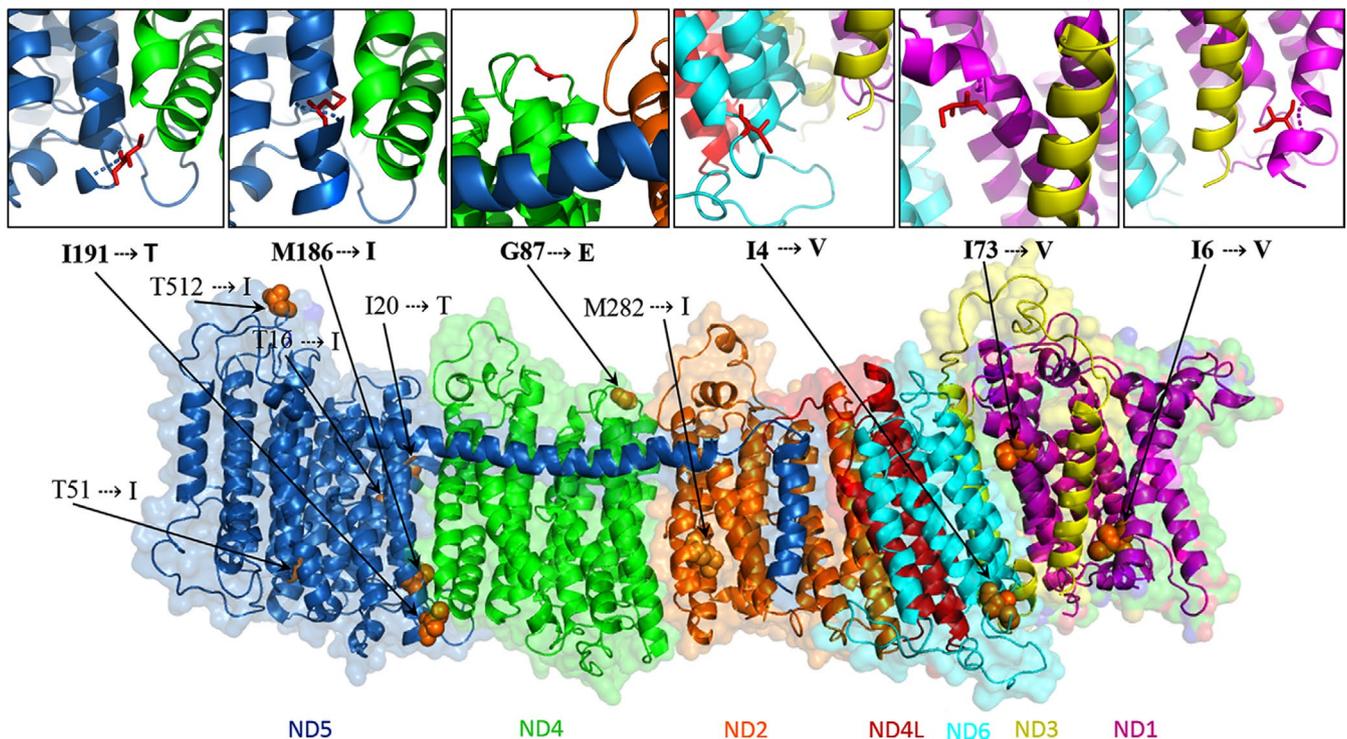


FIGURE 3 Atomic level structure of the predicted mitochondrial complex I of *M. primigenius*. Residues with amino acid substitutions are displayed as orange spheres. The potentially active sites are enlarged images shown as red sticks. Letters before the residues indicate the amino acids and their corresponding substitutions

Nonsynonymous mutations detected herein mostly influenced changes in amino acids ionization properties (Pk') (Table 2). Changes in ionization properties of the amino acids, including those of their side chains significantly influence protein's stability, solubility, and structural organization. At low ionic strength, most protein's solubility is relatively high, but it reduces with increase in the ionic strength (Cozzone, 2010). Similarly, the hydrophilicity/hydrophobicity of the side chains influences the physicochemical behavior of polypeptide chains including their foldability into three-dimensional structures (Cozzone, 2010). Therefore, changes in physicochemical properties in amino acids may have possibly affected the protein tertiary structures. This question was resolved by locating these residues in the protein structures.

3.3 | Amino acid substitutions in complex I

The homology protein structures of six out of seven OXPHOS complex I protein chains showed high % similarity (>63%) to the parent structure, suggesting a fair degree of structural similarities (Table 3). After homology modeling and side chain refinement of complex I structure, free loops that did not align with *O. aries* structure were excluded resulting to a final complex I structure of the woolly mammoth (Figure 3).

All the positively selected sites identified in complex I were mapped onto their corresponding subunits (Figure 3), to evaluate their locations and their relevance in protein structure and function. These sites were found in locations suggested not to be greatly

TABLE 4 Residue interactions with other protein subunits of complex I based on predictions from the protein structure

Gene (residue)	TM	Total TM	Binding residue	Interacting subunit (residues)	Relative accessible surface area (%)
ND1 (6)	1	8	+	ND3 (6,10,2,5,9)	0.1
ND1 (73)	2		+	ND6 (32)	32.7
ND2 (282)	7	8	-	-	0.0
ND4 (87)	Inside	12	-	-	0.0
ND5 (16)	1	14	-	-	50.9
ND5 (20)	1		-	-	38.6
ND5 (51)	2		-	-	15.6
ND5 (186)	5		+	ND4 (379)	0.5
ND5 (191)	Inside		+	ND2 (383,386,387,390)	27.5
ND5 (512)	Inside		-	-	66
ND6 (4)	1	5	+	1 (ND3)	54.5

Note: Inside indicates that the residue was located in the cytoplasmic matrix. TM represents the transmembrane.

TABLE 5 Amino acid differences on positively selected codons among the woolly mammoth mitolineages as inferred by TreeSAAP

Gene	Codon No.	<i>E. maximus</i>	<i>M. columbi</i>	<i>M. jefersonii</i>	<i>M. primigenius</i>				
					Clade I		Clade 2	Clade 3	
					C	D & E	A	B1	B2
ND1	6	V	V	V	V	V	I ⁽³⁾ -V ⁽¹⁾	V	V
ND1	73	V	V	V	V	V ⁽⁴⁸⁾ -I ⁽²⁾	I	V	V
ND2	282	M	I	I	I	I	M	M	M
ND4	87	G	G	G	G	G	G	E	E
ND5	16	T	T	T	T	T	I	T	T
ND5	20	T	T	T	I	T	T	I	I
ND5	51	T	T	T	T	T	I	T	T
ND5	186	I	I	I	I	I	M	I	I
ND5	191	I	I	I	T	T	I	I	I
ND5	512	T	I	I	I	I	T	T	T
ND6	4	I	I	I	I	I	V ⁽³⁾ -I ⁽¹⁾	I	I
Cyt <i>b</i>	189	I	I	I	I	I	I	M	M
Cyt <i>b</i>	300	I	I	I	I	I	I	V	V
Cyt <i>b</i>	371	I	M	M	M	M	I	I	I
ATP6	110	V	I	I	I	I	V	V	V
COX1	190	V	V	V	I	V	V	V	V

Note: The genus *M* represents *Mammuthus*. Numbers in the parenthesis indicate the number of individuals containing corresponding amino acids at each site.

Abbreviations: E, Glutamate; G, Glycine; I, Isoleucine; M, Methionine; T, Threonine; V, Valine.

conserved across species (Fiedorczuk et al., 2016). Mutations in five out of eleven residues in complex I were found in protein-protein binding sites where they interact with residues encoded by other subunits of complex I (Table 4). Amino acid changes at binding sites can influence binding properties of a residue including binding affinity, specificity, and foldability, which may affect protein-protein interaction during complex assembly and translocation process (Brender & Zhang, 2015). In addition, one substitution involved unrelated amino acids, which could likely cause changes in protein functions.

At residue ND1₆, isoleucine was substituted with valine. Here, all Clade I (haplogroups C, D, and E) and Clade 3 (haplogroups B1 & B2) along with M20 of Clade 2/A shared valine, while the rest of Clade 2/A shared isoleucine (Table 5). Chain ND1 is comprised of 8 transmembrane (TM) helices. Mutation in ND1₆ was found in TM1. ND1₆ was predicted to be a buried residue located in a protein-protein binding site, where it interacts with five other residues of subunit ND3. Both isoleucine and valine are aliphatic and hydrophobic, but valine is small (Taylor, 1986). Isoleucine and valine are not that

dissimilar in their biochemistry, therefore mutation in residue ND1₆ is not expected to cause direct changes in protein functions (Betts & Russell, 2003). But considering its location in an active binding site where it interacts with residues encoded by ND3 subunit, mutations at this site may affect the recognition of hydrophobic ligands (Betts & Russell, 2003) and protein–protein interactions during the assembly of OXPHOS complex I.

At residue ND1₇₃, substitution of isoleucine for valine was located in TM2. Here, all Clade 1 and Clade 3 shared valine except for samples NC007596 and DQ188829, both from Beringia which displayed isoleucine along with Clade 2/A. ND1₇₃ was predicted to be located near a possible proton-translocation pathway and binding site, where it interacts with a residue on the chain encoded by subunit ND6 (Table 4). In addition, ND1₇₃ interacts with residue ND6₃₀ that is near a functional residue ND6₃₂ suggested to be important in proton translocation (Sazanov, 2015). As such, mutations at this site could possibly affect the functional state of the proton-translocation pathway by coordinating the interacting residues.

Mutations in ND4₈₇ (glutamate to glycine) were located at the loop region and within the cytoplasmic matrix. This site is near the proposed proton-translocation pathway (Fiedorczuk et al., 2016; Sazanov, 2015). Here, all Clade 1 and Clade 2/A displayed glutamate, while Clade 3 showed glycine. Glutamate is charged and polar, while glycine is hydrophobic and small. Glutamate can be involved in salt bridges and play a role in catalytic sites of proteins (Betts & Russell, 2003), whereas glycine can play a functional role by binding to phosphates of the ATP molecule (Schulze-Gahmen, Bondt, & Kim, 1996). Considering the location and physicochemical dissimilarities of amino acids change at ND4₈₇, mutation at this site might have caused possible changes in the proton-translocation channel and influenced its efficiency.

At ND5, which contains fourteen TM Helices, amino acid replacements observed at ND5₁₆ (isoleucine to threonine) and ND2₀ (threonine to isoleucine), were both located at TM1. Residue ND5₁₆ was a surface residue, while ND2₀ was near the surface. All Clade 1 shared threonine at ND5₁₆, while Clade 2/A had isoleucine. On the other hand, isoleucine at ND2₀ was shared by all Clade 1 while threonine was present in Clade 2/A. Isoleucine is hydrophobic and aliphatic, with threonine being hydrophobic, polar, and small. Unlike isoleucine, threonine contains a fairly reactive hydroxyl group, which can form hydrogen bonds with various substrates (Betts & Russell, 2003). These two residues (ND5₁₆, ND2₀) were not located in binding sites, but considering their location on/near the surface, the nature of amino acid substitution, and the role of the mutations in driving radical changes of six physicochemical properties, it is highly likely that they had a functional role.

Mutations in residues ND5₁₈₆ and ND5₁₉₁ were both located at protein–protein binding sites, which interacts with one residue on chain encoded by ND4, and four residues of chain encoded by ND2 subunits, respectively (Table 5). Subunits ND2, ND4, and ND5 are suggested to be proton-pumping machines that are related to the Na⁺/H⁺ antiporters of the multiple resistance and PH (Mrp) family (Fonseca et al., 2008). Interestingly, substitution in ND5₁₈₆

and ND5₁₉₁ were located near a possible Na⁺ translocation channel, which has been proposed in *Escherichia coli* complex I (Sazanov, 2015). Previous evidence suggests that Na⁺ could be playing a direct role in catalytic mechanisms of complex 1 (Batista, Marreiros, & Pereira, 2012). Specifically, site ND5₁₈₆ was predicted to be a buried residue located in the TM5. Here, all Clade 1 and Clade 3 shared isoleucine while Clade 2/A displayed methionine. Both methionine and isoleucine are hydrophobic, but isoleucine is aliphatic (Betts & Russell, 2003; Taylor, 1986). Meanwhile, the mutation in ND5₁₉₁ was located in cytoplasmic region, where threonine was replaced with isoleucine. At this site, all Clade 1 had threonine, while Clade 2/A and Clade 3 shared isoleucine (Table 5). Taking into consideration that ND5₁₈₆ and ND5₁₉₁ are binding sites for other residues in chains encoded by ND4 and ND2 subunits, respectively, coupled with their location near a possible Na⁺ translocation channel, mutations at these residues might have likely influenced protein functions including catalytic mechanisms in complex I. Mutations affecting H⁺/Na⁺ transport activity has been found in dolphins, although their key role in adaptive selection could not be determined (Caballero, Duchêne, Garavito, Slikas, & Baker, 2015). Lastly, substitution ND6₄ (isoleucine for valine) was located in TM1. ND6 has 5 TM's in total. At ND6₄, which is located at the loop region, all Clade 1, Clade 3, and M20 of Clade 2/A displayed isoleucine, while the rest of Clade 2/A shared valine (Table 5). This substitution was found at the surface and binding site for a functional residue (site 1, at the N-terminus of ND3), which makes it more likely to influence the overall protein structure and efficiency of the proton-translocation channel which ND6 forms with ND1 and ND4L subunits.

The amino acids substitutions identified above are not that dissimilar in biochemistry, but considering their location, they may likely drive conformational changes in the proteins and the efficiency of the translocation processes including OXPHOS process.

3.4 | Amino acid substitutions of complex III, IV, and V

We also observed significant amino acid substitutions in cyt *b*, COX1, and ATP6 subunits, which are membrane-bound proteins found in mitochondrial complex III, IV, and V, respectively. Due to the difficulty in crystalizing membrane-bound proteins, there is still limited structural information about the subunits such as for F0 proton channel in ATP6 important in OXPHOS process (Weber & Senior, 1997) and the hemes' dynamics in complex III which are important in electron transport chain (Kim, Khalimonchuk, Smith, & Winge, 2012). Therefore, we did not perform additional structural analysis of the aforementioned protein chains, but rather examined their biochemical differences.

Amino acid replacements in cyt *b* were found at residues 189, 300, and 371. At site cyt *b*₁₈₉, methionine was substituted with isoleucine, where isoleucine was shared by Clade 1 and 2/A, while methionine was private to Clade 3 (Table 5). Both methionine and isoleucine are hydrophobic, but isoleucine is aliphatic. In residue cyt *b*₃₀₀, valine was substituted with isoleucine. Here, all Clade 1 and

Clade 2/A shared isoleucine while Clade 3 haplogroups displayed valine. Both of these amino acids are hydrophobic but valine is small. At *cyt b*₃₇₁, isoleucine was replaced by methionine, whose properties have been mentioned above. At this site, Clade 1 displayed methionine while Clade 2/A and Clade 3 shared isoleucine. With respect to COX1, mutations were found at site 190. Here, valine was replaced with isoleucine, where all Clade 1 shared isoleucine, while the rest of the clades shared valine (Table 4). Again, both of these amino acids are hydrophobic with valine being small. Lastly, at ATP6₁₁₀, valine was substituted with isoleucine. At this site, all Clade 1 shared isoleucine while the rest of the clades shared valine. Generally, mutations in the *cyt b*, COX1, and ATP6 residues involved closely related amino acid residues, thus, they were not expected to drive significant changes in protein structure and function (Betts & Russell, 2003).

4 | DISCUSSION

Woolly mammoth comprised of endemic mitolineages that is Clade 2/A and Clade 1 haplogroup C that were exclusively Asian and North American in distribution, respectively, whereas, Clade 1 (haplogroups D & E) and Clade 3 were predominantly distributed in Asia and Europe, respectively. Such geographic pattern suggests a high level of intercontinental differentiation (Debruyne et al., 2008) and survival in varied environments. In order to detect positive selection in the mitogenome in woolly mammoths, physicochemical changes among codons in a phylogenetic tree were analyzed coupled with inferences of the potential effects of amino acid substitutions on the protein structure and function.

At the biochemical level, we found sixteen sites with strong positive selection leading to changes in amino acids physicochemical properties in ND1, ND2, ND4, ND5, ND6, CYTB, ATP6, and COX1 (Table 2). All the thirteen observed amino acid substitutions in the aforementioned genes affected the ionization properties of the amino acids, while five of them also led to changes in five other properties influencing hydrophilicity or compactness in the region surrounding the residues under mutation. Changes in these functional properties may compromise protein's stability, solubility, structural organization, and foldability. These findings showed slight disparity with the previous studies in woolly mammoths, where positive selection signal in complex I genes was not detected (Gilbert et al., 2008; Pečnerová et al., 2017), possibly because we used expanded dataset for divergent clades with a larger geographic coverage, thus providing a larger number of variable codons. We also drew inferences from TreeSAAP algorithm that is highly sensitive to even the most conservative proteins.

Our findings suggest that positive selection might have had a preferential effect on residues in NADH dehydrogenase (ND) complex (complex I). The residue mutations in complex I subunits were found to occur mainly in the transmembrane helices and regions that are not greatly conserved (Fiedorczuk et al., 2016). With respect to ND1 subunit, mutations were found in residues located at

protein–protein binding sites. ND1 is fundamental in the assembly of complex I, since they serve as electron transport conduits, from ubiquinone to the binding pocket in the Q-Module, thus, changes in ND1 can affect the assembly of the entire complex (Cardol, Matagne, & Remacle, 2002). Moreover, mutations in binding sites may affect protein–protein interaction during complex assembly and translocation process (Brender & Zhang, 2015). Mutations in ND1 have been previously implicated in influencing metabolic efficiency in varied species including salmon (Consuegra, John, Verspoor, & Leaniz, 2015) and African elephants (Finch, Zhao, Korkin, Frederick, & Eggert, 2014). Finch et al., (2014) argue that mutations in ND1, ND4, ND5, ND6, and ATP6 genes in the African elephants are an indicative of adaptive evolution. Considering that mutations in ND1 detected in this study involved residues found at protein–protein binding sites where they interacted with residues in other subunits (Table 4), this may be a reflection of their potential importance in protein function, most likely in metabolic processes.

Other mutations detected in the current study were at residues in ND2, ND4, ND5, and ND6 as shown in Table 2. In mammalian mitogenome, ND2, ND4, and ND5 genes are regarded as the real proton pumps related to Na⁺/H⁺ antiporters of the Mrp family in OXPHOS complex I (Fonseca et al., 2008). These three subunits are connected by the arm of ND5, which allows for well-coordinated shifts in proton pumping that consequently drive ATP production (Hunte, Zickermann, & Brandt, 2010). Changes (e.g., biochemical changes) in these subunits may hinder or improve proton-pumping process (Fonseca et al., 2008). Mutations at ND4₈₇ (found in all lineages except Clade 3), ND5₁₆, ND5₁₈₆ (found in all lineages except Clade 2/A), ND5₁₉₁ (private to Clade 1 haplogroup C, D and E), and ND6₄ (private to clade except for one sample) were all located in protein–protein interacting sites or/and translocation channels where they were likely to influence biochemical processes (Table 4). Interestingly, mutations in some of the aforementioned residues coincided with those reported elsewhere; ND5₁₆ mutation is suggested to have a role in functional metabolic differences in *Tachycineta* swallow species inhabiting environments with latitudinal variations (Stager et al., 2014). Mutations at site ND5₁₆ (of amphibians and coelacanths) and ND5₂₀ (of amphibians, lungfishes, and coelacanths) are suggested to be linked to evolutionary adaptation to high atmospheric oxygen during cladogenesis (George & Blicek, 2011). Finch et al. (2014) in their study on African savanna and forest elephants also proposed that positive selection in complex I residues including ND5₂₀ is indicative of ecological adaptations. With respect to site ND6₄, mutations in residues adjacent to this site have been linked to adaptive selection, where amino acid replacements in ND6₆ is suggested to influence efficiency of electron transfer under cold and hypoxic conditions in cephalopods (Almeida et al., 2015), while in rodents, ND6₅ and ND6₆ mutations are linked with adaptation to hypoxic conditions (Tomasco & Lessa, 2011). This evidence points to adaptive variations to oxygen availability and metabolic processes, which may be associated with possible metabolic adaptation of

woolly mammoths clades in cold high-latitude environments that are characterized with high oxygen pressure. This hypothesis may also be correlated with previous findings on the adaptation of woolly mammoth hemoglobin in maintaining stable O₂ affinity in extremely cold environments (Campbell et al., 2010).

In our analyses, significant mutations were also detected in *cyt b*, ATP6, and COX1. *Cyt b* is involved in catalyses of the reversible transfer of electrons from ubiquinol to the cytochrome c coupled with proton translocation against the proton gradient (Q-cycle). This proton gradient is utilized by ATP synthase to convert adenosine diphosphate (ADP) into adenosine triphosphate (ATP) (Saraste, 1999). Based on our biochemical analyses, we did not find clear evidence for functional relevancy of the mutations in *cyt b*, ATP6, and COX1, and their involvement in adaptive selection in woolly mammoths. However, we have taken note of previous findings where mutations in these sites were reported. In Gilbert, Drautz (Gilbert et al., 2008) study on woolly mammoths, mutations at *cyt b*₃₇₁ and COX1₁₉₀ were not suggested to have functional importance in metabolic processes. On the other hand, mutations at *cyt b*₃₀₀ (in amphibians and lungfishes) and ATP6₁₁₀ (in amphibians and coelacanths) have been linked with adaptation to increased oxygen levels (George & Blicek, 2011). Taking into consideration of a combination of this evidence with our results, where mutations in *cyt b*₃₀₀ (shared by clade 1 and 2/A) and ATP6₁₁₀ (private to clade 1) were only present in lineages that predominantly existed in persistently cold and dry environments in Northern Siberia and North America (Lister, Sher, Essen, & Wei, 2005; Muhs, Ager, Been, Bradbury, & Dean, 2003; Szpak et al., 2010), this raises a possibility that these substitutions might have contributed to a regulatory-driven adaptation.

Given the expanded mitogenome dataset and robust inference methods employed in this study, we are confident that this work evaluated positive selection signals in mitochondrial protein-coding genes across woolly mammoth's phylogeny. In conclusion, our findings provide compelling evidence for positive selection in eleven promising candidate residues in the OXPHOS genes, which suggests that these mutations might have played a role in metabolic adaptations among the lineages to their different environments. However, these inferences can be further tested in the future to provide more insights into the effects of these mutations in other systems and link their relevancy to shaping woolly mammoth's mitochondrial evolution.

ACKNOWLEDGMENTS

This study was supported by the Fundamental Research Funds for the Central Universities of China (2017), Wildlife Conservation and Management Project of State Forestry Administration of China (2016), the Guangdong Provincial Key Laboratory of Genome Read and Write (grant no. 2017B030301011), and Guangdong Provincial Academician Workstation of BGI Synthetic Genomics, BGI-Shenzhen, Guangdong, China. The authors are grateful to Mr. Yue Ma, Mr. Wei Zhang, Miss Qing Wang, Miss Li Liu and Mr. Xin Mu at Northeast Forestry University, China, for their discussions and recommendations.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR'S CONTRIBUTIONS

JNN and YCX designed the study. TML prepared the complete mitochondrial genome sequences from whole genome sequences. JNN and TDD performed analyses, and JNN drafted the manuscript. LZ and AKA provided valuable suggestions during data analyses. All authors contributed to and approved the final manuscript.

DATA ACCESSIBILITY

Mitochondrial genome sequences generated in this study have been stored in China National GeneBank (CNGB) Nucleotide Sequence Archive (CNSA: <https://db.cngb.org/cnsa>; Accession No. CNP0000277).

ORCID

Ahmed Khalid Ahmed  <https://orcid.org/0000-0001-8301-8510>

Yan Chun Xu  <https://orcid.org/0000-0003-0397-4292>

REFERENCES

- Almeida, D., Maldonado, E., Vasconcelos, V., & Antunes, A. (2015). Adaptation of the mitochondrial genome in Cephalopods: Enhancing proton translocation channels and the subunit interactions. *PLoS ONE*, 10(8), e0135405. <https://doi.org/10.1371/journal.pone.0135405>
- Bai, Y., Shakeley, R. M., & Attardi, G. (2000). Tight control of respiration by NADH dehydrogenase ND5 subunit gene expression in mouse mitochondria. *Molecular and Cellular Biology*, 20(3), 805–815. <https://doi.org/10.1128/MCB.20.3.805-815.2000>
- Ballard, J., & Melvin, R. (2010). Linking the mitochondrial genotype to the organismal phenotype. *Molecular Ecology*, 19(8), 1523–1539. <https://doi.org/10.1111/j.1365-294X.2010.04594.x>
- Barnes, I., Shapiro, B., Lister, A., Kuznetsova, T., Sher, A., Guthrie, D., & Thomas, M. G. (2007). Genetic structure and extinction of the woolly mammoth, *Mammuthus primigenius*. *Current Biology*, 17(12), 1072–1075. <https://doi.org/10.1016/j.cub.2007.05.035>
- Batista, A. P., Marreiros, B. C., & Pereira, M. M. (2012). The role of proton and sodium ions in energy transduction by respiratory complex I. *IUBMB Life*, 64(6), 492–498.
- Betts, M. J., & Russell, R. B. (2003). Amino acid properties and consequences of substitutions. In M. R. Barnes, & I. C. Gray (Eds.), *Bioinformatics for geneticists* (pp. 289–316). Hoboken, NJ: John Wiley & Sons.
- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3), 345–370. <https://doi.org/10.1007/BF02294361>
- Breder, J. R., & Zhang, Y. (2015). Predicting the effect of mutations on protein-protein binding interactions through structure-based interface profiles. *PLoS Computational Biology*, 11(10), e1004494. <https://doi.org/10.1371/journal.pcbi.1004494>
- Caballero, S., Duchêne, S., Garavito, M. F., Slikas, B., & Baker, C. S. (2015). Initial evidence for adaptive selection on the NADH subunit two of freshwater dolphins by analyses of mitochondrial genomes. *PLoS ONE*, 10(5), e0123543. <https://doi.org/10.1371/journal.pone.0123543>

- Campbell, K. L., Roberts, J. E. E., Watson, L. N., Stetefeld, J., Sloan, A. M., Signore, A. V., ... Cooper, A. (2010). Substitutions in woolly mammoth hemoglobin confer biochemical properties adaptive for cold tolerance. *Nature Genetics*, 42, 536. <https://doi.org/10.1038/ng.574>
- Cardol, P., Matagne, R. F., & Remacle, C. (2002). Impact of mutations affecting ND mitochondria-encoded subunits on the activity and assembly of complex I in *Chlamydomonas*. Implication for the structural organization of the enzyme. *Journal of Molecular Biology*, 319(5), 1211–1221. [https://doi.org/10.1016/S0022-2836\(02\)00407-2](https://doi.org/10.1016/S0022-2836(02)00407-2)
- Chang, D., Knapp, M., Enk, J., Lippold, S., Kircher, M., Lister, A., ... Shapiro, B. (2017). The evolutionary and phylogeographic history of woolly mammoths: A comprehensive mitogenomic analysis. *Scientific Reports*, 7, 44585. <https://doi.org/10.1038/srep44585>
- Consuegra, S., John, E., Verspoor, E., & De Leaniz, C. G. (2015). Patterns of natural selection acting on the mitochondrial genome of a locally adapted fish species. *Genetics Selection Evolution*, 47(1), 58. <https://doi.org/10.1186/s12711-015-0138-0>
- Cozzone, A. J. (2010). Proteins: Fundamental chemical properties. eLS.
- da Fonseca, R. R., Johnson, W. E., O'Brien, S. J., Ramos, M. J., & Antunes, A. (2008). The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics*, 9(1), 119. <https://doi.org/10.1186/1471-2164-9-119>
- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, 9, 772. <https://doi.org/10.1038/nmeth.2109>
- Debruyne, R., Chu, G., King, C. E., Bos, K., Kuch, M., Schwarz, C., ... Poinar, H. N. (2008). Out of America: Ancient DNA evidence for a new world origin of late quaternary woolly mammoths. *Current Biology*, 18(17), 1320–1326. <https://doi.org/10.1016/j.cub.2008.07.061>
- DeLano, W. L. (2002). The PyMOL molecular graphics system. Retrieved from <http://pymol.org>
- Enk, J., Devault, A., Widga, C., Saunders, J., Szpak, P., Southon, J., ... Poinar, H. (2016). Mammuthus population dynamics in late Pleistocene North America: Divergence, phylogeography, and introgression. *Frontiers in Ecology and Evolution*, 4, 42. <https://doi.org/10.3389/fevo.2016.00042>
- Fellows Yates, J. A., Drucker, D. G., Reiter, E., Heumos, S., Welker, F., Münzel, S. C., ... Krause, J. (2017). Central European woolly mammoth population dynamics: insights from Late Pleistocene Mitochondrial Genomes. *Scientific Reports*, 7(1), 17714. <https://doi.org/10.1038/s41598-017-17723-1>
- Fiedorczuk, K., Letts, J. A., Degliesposti, G., Kaszuba, K., Skehel, M., & Sazanov, L. A. (2016). Atomic structure of the entire mammalian mitochondrial complex I. *Nature*, 538(7625), 406. <https://doi.org/10.1038/nature19794>
- Finch, T. M., Zhao, N., Korkin, D., Frederick, K. H., & Eggert, L. S. (2014). Evidence of positive selection in mitochondrial complexes I and V of the African Elephant. *PLoS ONE*, 9(4), e92587. <https://doi.org/10.1371/journal.pone.0092587>
- Fraczkiewicz, R., & Braun, W. (1998). Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *Journal of Computational Chemistry*, 19(3), 319–333. [https://doi.org/10.1002/\(SICI\)1096-987X\(199802\)19:3<319:AID-JCC6>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1096-987X(199802)19:3<319:AID-JCC6>3.0.CO;2-W)
- George, D., & Blicek, A. (2011). Rise of the earliest tetrapods: An early Devonian origin from marine environment. *PLoS ONE*, 6(7), e22136. <https://doi.org/10.1371/journal.pone.0022136>
- Gilbert, M. T. P., Drautz, D. I., Lesk, A. M., Ho, S. Y. W., Qi, J., Ratan, A., ... Schuster, S. C. (2008). Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 105(24), 8327–8332. <https://doi.org/10.1073/pnas.0802315105>
- Gilbert, M. T. P., Tomsho, L. P., Rendulic, S., Packard, M., Drautz, D. I., Sher, A., ... Schuster, S. C. (2007). Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science*, 317(5846), 1927–1930. <https://doi.org/10.1126/science.1146971>
- Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., & Ben-Tal, N. (2003). ConSurf: Identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, 19(1), 163–164. <https://doi.org/10.1093/bioinformatics/19.1.163>
- Gu, J., Wu, M., Guo, R., Yan, K., Lei, J., Gao, N., & Yang, M. (2016). The architecture of the mammalian respirasome. *Nature*, 537(7622), 639. <https://doi.org/10.1038/nature19359>
- Hunte, C., Zickermann, V., & Brandt, U. (2010). Functional modules and structural basis of conformational coupling in mitochondrial complex I. *Science*, 329(5990), 448–451. <https://doi.org/10.1126/science.1191046>
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Kim, H. J., Khalimonchuk, O., Smith, P. M., & Winge, D. R. (2012). Structure, function, and assembly of heme centers in mitochondrial respiratory complexes. *Biochimica Et Biophysica Acta (BBA) - Molecular Cell Research*, 1823(9), 1604–1616. <https://doi.org/10.1016/j.bbamcr.2012.04.008>
- Krause, J., Dear, P. H., Pollack, J. L., Slatkin, M., Spriggs, H., Barnes, I., ... Hofreiter, M. (2006). Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. *Nature*, 439(7077), 724. <https://doi.org/10.1038/nature04432>
- Krogh, A., Larsson, B., vonHeijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. 2001(0022–2836 (Print)).
- Lecocq, T., Vereecken, N. J., Michez, D., Dellicour, S., Lhomme, P., Valterová, I., ... Rasmont, P. (2013). Patterns of genetic and reproductive traits differentiation in mainland vs. Corsican populations of bumblebees. *PLoS ONE*, 8(6), e65642. <https://doi.org/10.1371/journal.pone.0065642>
- Levin, L., Blumberg, A., Barshad, G., & Mishmar, D. (2014). Mito-nuclear co-evolution: The positive and negative sides of functional ancient mutations. *Frontiers in Genetics*, 5, 448. <https://doi.org/10.3389/fgene.2014.00448>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Lindgreen, S. (2012). AdapterRemoval: Easy cleaning of next-generation sequencing reads. *BMC Research Notes*, 5(1), 337. <https://doi.org/10.1186/1756-0500-5-337>
- Lister, A., & Bahn, P. (2007). *Mammoths: Giants of the Ice Age* (revised edition). London, UK: Frances Lincoln.
- Lister, A. M., Sher, A. V., van Essen, H., & Wei, G. (2005). The pattern and process of mammoth evolution in Eurasia. *Quaternary International*, 126, 49–64. <https://doi.org/10.1016/j.quaint.2004.04.014>
- MacDonald, G. M., Beilman, D. W., Kuzmin, Y. V., Orlova, L. A., Kremenetski, K. V., Shapiro, B., ... Van Valkenburgh, B. (2012). Pattern of extinction of the woolly mammoth in Beringia. *Nature Communications*, 3, 893. <https://doi.org/10.1038/ncomms1881>
- McClellan, D. A., & Ellison, D. D. (2010). Assessing and improving the accuracy of detecting protein adaptation with the TreeSAAP analytical software. *International Journal of Bioinformatics Research and Applications*, 6(2), 120–133. <https://doi.org/10.1504/IJBRA.2010.032116>
- Muhs, D. R., Ager, T. A., Been, J., Bradbury, J. P., & Dean, W. E. (2003). A late Quaternary record of eolian silt deposition in a maar lake, St. Michael Island, Western Alaska. *Quaternary Research*, 60(1), 110–122. [https://doi.org/10.1016/S0033-5894\(03\)00062-0](https://doi.org/10.1016/S0033-5894(03)00062-0)

- Palkopoulou, E., Dalen, L., Lister, A. M., Vartanyan, S., Sablin, M., Sher, A., ... Thomas, J. A. (2013). Holarctic genetic structure and range dynamics in the woolly mammoth. *Proceedings of the Royal Society B: Biological Sciences*, 280(1770), 20131910. <https://doi.org/10.1098/rspb.2013.1910>
- Palkopoulou, E., Mallick, S., Skoglund, P., Enk, J., Rohland, N., Li, H., ... Dalén, L. (2015). Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Current Biology*, 25(10), 1395–1400. <https://doi.org/10.1016/j.cub.2015.04.007>
- Pečnerová, P., Palkopoulou, E., Wheat, C. W., Skoglund, P., Vartanyan, S., Tikhonov, A., ... Dalén, L. (2017). Mitogenome evolution in the last surviving woolly mammoth population reveals neutral and functional consequences of small population size. *Evolution Letters*, 1(6), 292–303. <https://doi.org/10.1002/evl3.33>
- Petřek, M., Otyepka, M., Banáš, P., Košíňová, P., Koča, J., & Damborský, J. (2006). CAVER: A new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics*, 7(1), 316.
- Petterson, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. <https://doi.org/10.1002/jcc.20084>
- Rand, D. M., Haney, R. A., & Fry, A. J. (2004). Cytonuclear coevolution: The genomics of cooperation. *Trends in Ecology & Evolution*, 19(12), 645–653. <https://doi.org/10.1016/j.tree.2004.10.003>
- Rogaev, E. I., Moliaka, Y. K., Malyarchuk, B. A., Kondrashov, F. A., Derenko, M. V., Chumakov, I., & Grigorenko, A. P. (2006). Complete mitochondrial genome and phylogeny of Pleistocene mammoth *Mammuthus primigenius*. *PLoS Biology*, 4(3), e73. <https://doi.org/10.1371/journal.pbio.0040073>
- Rohland, N., Malaspina, A.-S., Pollack, J. L., Slatkin, M., Matheus, P., & Hofreiter, M. (2007). Proboscidean mitogenomics: Chronology and mode of elephant evolution using mastodon as outgroup. *PLoS Biology*, 5(8), e207. <https://doi.org/10.1371/journal.pbio.0050207>
- Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12), 1572–1574. <https://doi.org/10.1093/bioinformatics/btg180>
- Saraste, M. (1999). Oxidative phosphorylation at the fin de siècle. *Science*, 283(5407), 1488–1493.
- Sazanov, L. A. (2015). A giant molecular proton pump: Structure and mechanism of respiratory complex I. *Nature Reviews Molecular Cell Biology*, 16(6), 375. <https://doi.org/10.1038/nrm3997>
- Schulze-Gahmen, U., De Bondt, H. L., & Kim, S.-H. (1996). High-resolution crystal structures of human cyclin-dependent kinase 2 with and without ATP: Bound waters and natural ligand as guides for inhibitor design. *Journal of Medicinal Chemistry*, 39(23), 4540–4546.
- Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., & Serrano, L. (2005). The FoldX web server: An online force field. *Nucleic Acids Research*, 33(suppl_2), W382–W388. <https://doi.org/10.1093/nar/gki387>
- Smith, S. D., Kawash, J. K., Karaiskos, S., Biluck, I., & Grigoriev, A. (2017). Evolutionary adaptation revealed by comparative genome analysis of woolly mammoths and elephants. *DNA Research*, 24(4), 359–369. <https://doi.org/10.1093/dnares/dsx007>
- Stager, M., Cerasale, D. J., Dor, R., Winkler, D. W., & Cheviron, Z. A. (2014). Signatures of natural selection in the mitochondrial genomes of Tachycineta swallows and their implications for latitudinal patterns of the 'pace of life'. *Gene*, 546(1), 104–111. <https://doi.org/10.1016/j.gene.2014.05.019>
- Szpak, P., Gröcke, D. R., Debruyne, R., MacPhee, R. D. E., Guthrie, R. D., Froese, D., ... Poinar, H. N. (2010). Regional differences in bone collagen $\delta^{13}C$ and $\delta^{15}N$ of Pleistocene mammoths: Implications for paleoecology of the mammoth steppe. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 286(1–2), 88–96. <https://doi.org/10.1016/j.palaeo.2009.12.009>
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, 28(10), 2731–2739. <https://doi.org/10.1093/molbev/msr121>
- Taylor, W. R. (1986). The classification of amino acid conservation. *Journal of Theoretical Biology*, 119(2), 205–218. [https://doi.org/10.1016/S0022-5193\(86\)80075-3](https://doi.org/10.1016/S0022-5193(86)80075-3)
- Tomasco, I. H., & Lessa, E. P. (2011). The evolution of mitochondrial genomes in subterranean caviomorph rodents: Adaptation against a background of purifying selection. *Molecular Phylogenetics and Evolution*, 61(1), 64–70. <https://doi.org/10.1016/j.ympev.2011.06.014>
- Vangone, A., & Bonvin, A. M. J. J. (2015). Contacts-based prediction of binding affinity in protein–protein complexes. *eLife*, 4, e07454. <https://doi.org/10.7554/eLife.07454>
- Wallace, D. C. (2007). Why do we still have a maternally inherited mitochondrial DNA? Insights from evolutionary medicine. *Annual Review of Biochemistry*, 76, 781–821. <https://doi.org/10.1146/annurev.biochem.76.081205.150955>
- Weber, J., & Senior, A. E. (1997). Catalytic mechanism of F1-ATPase. *Biochimica Et Biophysica Acta (BBA)-Bioenergetics*, 1319(1), 19–58. [https://doi.org/10.1016/S0005-2728\(96\)00121-1](https://doi.org/10.1016/S0005-2728(96)00121-1)
- Woolley, S., Johnson, J., Smith, M. J., Crandall, K. A., & McClellan, D. A. (2003). TreeSAAP: Selection on amino acid properties using phylogenetic trees. *Bioinformatics*, 19(5), 671–672. <https://doi.org/10.1093/bioinformatics/btg043>
- Wu, M., Gu, J., Guo, R., Huang, Y., & Yang, M. (2016). Structure of mammalian respiratory supercomplex I1III2IV1. *Cell*, 167(6), 1598–1609. e10. <https://doi.org/10.1016/j.cell.2016.11.012>
- Xue, L. C., Rodrigues, J. P., Kastriitis, P. L., Bonvin, A. M., & Vangone, A. (2016). PRODIGY: A web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics*, 32(23), 3676–3678. <https://doi.org/10.1093/bioinformatics/btw514>
- Yang, M. A., Gao, X., Theunert, C., Tong, H., Aximu-Petri, A., Nickel, B., ... Fu, Q. (2017). 40,000-year-old individual from Asia provides insight into early population structure in Eurasia. *Current Biology*, 27(20), 3202–3208.e9. <https://doi.org/10.1016/j.cub.2017.09.030>
- Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution*, 15(5), 568–573. <https://doi.org/10.1093/oxfordjournals.molbev.a025957>
- Yu, L., Wang, X., Ting, N., & Zhang, Y. (2011). Mitogenomic analysis of Chinese snub-nosed monkeys: Evidence of positive selection in NADH dehydrogenase genes in high-altitude adaptation. *Mitochondrion*, 11(3), 497–503. <https://doi.org/10.1016/j.mito.2011.01.004>
- Zhu, J., Vinokumar, K. R., & Hirst, J. (2016). Structure of mammalian respiratory complex I. *Nature*, 536(7616), 354. <https://doi.org/10.1038/nature19095>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Ngatia JN, Lan TM, Dinh TD, Zhang L, Ahmed AK, Xu YC. Signals of positive selection in mitochondrial protein-coding genes of woolly mammoth: Adaptation to extreme environments? *Ecol Evol*. 2019;9:6821–6832. <https://doi.org/10.1002/ece3.5250>